A Computer System

This invention relates to a computer system, in particular to a computer system having a plurality of components which can be initialised.

Despite advances in computer technology, the components of a computer system can have deficiencies which cause faults to develop over time. For example, many applications, such as those written in the language "C" contain memory leaks, wherein the application does not release memory after it has finished using it. Such faults can cause a component to hang or crash, thereby hindering the efficient operation of the system. It is know to cure or prevent a fault by initialising the faulty or potentially faulty component, for example by replacing at least some of the software from the component, by re-starting or re-booting the component, or by re-setting one or more control parameters of that component. However, the initialisation process will often disrupt the normal operation of the component, at least during the period in which initialisation is taking place. During this period, other components may be required to perform an additional number of tasks to compensate for the tasks that are not being performed by the disrupted component. Such a re-distribution of tasks may reduce the efficiency of the system as a whole.

According to one aspect of the present invention, there is provided a computer system having a plurality of components that can be initialised, wherein each component is configured to produce status data from which the level of need for that component to be initialised can be inferred, the status data having a predetermine level of need associable therewith and wherein at least one component is configured to: receive status data from other components; make a comparison using the status data received from respective components; in dependence on the comparison, select one or more components for initialisation; and, issue initialisation instructions to the selected component(s).

By making a comparison between the status data from different components, the relative need for different components to be initialised can be evaluated, allowing a component with a high need for initialisation to be selected over a component with a lower need. Thus, rather than considering in isolation whether a component needs to be initialised, the present invention allows the initialisation of components to be prioritised, thereby reducing the amount of disruption to the system as a whole.

The components may be hardware components, or alternatively, the components may be software components, running on at least one (hardware) computer device. In one embodiment, the components are software components each running on a respective computer device, the initialisation of a component causing the device on which it runs to be re-booted or otherwise initialised.

The status data received from a component may include a plurality of status values, such as the amount of free memory in a memory location, the frequency with which a processor is accessing a memory location, or other historical or current data which can be used to determine the need for initialisation of the component.

However, to simplify the processing to be carried out by a component receiving the status data, the status data will preferably be in the form of an initialisation parameter. This will reduce the need for a component to evaluate and/or cross reference different data originating from the same component in order to infer the need for that component to be initialised. Furthermore, if the need for a component to be initialised can be expressed simply in terms of a variable, such as the elapsed time since the last initialisation, the respective values of that variable for different components can be directly compared against one another.

The level of need for a component to be initialised can be positive or negative. Thus an initialisation parameter may indicate the importance or urgency for a component to be initialised, but an initialisation parameter may alternatively indicate the importance for a component not to be initialised, for example if that component is carrying out an essential task.

Preferably, each component will be configured to execute an initialisation routine when the initialisation parameter for that component reaches a threshold value, in which case that component will behave as an initialising component, sending a request message to other (recipient) components, the request message requesting respective initialisation parameters from the recipient components. Hence, when the need for a component to be initialised becomes sufficiently high, that component can at least attempt to find out the need for other components to be initialised. This allows the amount of information which a component needs to retain about the state of other components to be reduced, since a component can when required obtain such information by transmitting a request message.

Furthermore, if each component can request initialisation parameters from other components, the need for the initialisation requirements of different components to be stored at a central location is reduced.

The threshold value for each component may be the same or different, and may be set by a system administrator in dependence on the importance or the number of tasks performed by a component.

The initialisation routine carried out by an initialising component preferably includes the further steps of: comparing the initialisation parameters received from other components with the initialisation parameter for the initialising component; and, in dependence on the comparison, making a self-initialisation decision. In this way, each component can take into account the need of other components to initialise before taking a self-initialisation decision.

In a preferred embodiment, each component is configured to select components for initialisation, and to issue initialisation instructions to the selected components. Thus, in this embodiment, even if one component is unable, due to a fault or otherwise, to act as an initiating component and select components for initialisation, another component of the system will eventually be triggered to act as an initiating device (for example because the time since the last initialisation of that component has exceeded a threshold value). This fault tolerance will allow the system to maintain itself efficiently, since the task of selecting which components to initialise can in effect be distributed across the system.

The status data associable with a value of need data will preferably be predetermined such that the data arriving at a component from other components will have a level of need associable or associated therewith when it arrives. Thus, the association between status data and the need for initialisation will preferably be predetermined such that it is possible to infer an indication of the level of need for a component to be initialised before a comparison is made between the status data from the different components. The comparison between the status data from different components can then be used to evaluate the relative need of two or more components to be initialised, relative to one another.

A receiving component may store a mapping or a functional relationship or other association between the values which status data can take on the one hand and levels of need on the other hand. For example association may simply be a one-to-one mapping between the level of need and values of status data, e.g., the higher the level of need, the greater the magnitude of the status data value. However, it will preferably be possible to alter a relationship between the value of need and the status data, for example in response to a changing system environment. Thus (for any given relationship between status data and levels of need) the status data may be considered to be indicative of need.

Preferably, the status data will be able to take one or more values within a (possibly discrete) range of values. For example, the status data may take one (or more) of (at least) three values, indicative of high, medium or low initialisation need. The status data from different components can then be at least partially ordered in dependence on (or accordance with) the relative value of the status data from at least some of the different components.

The invention will now be further described, by way of example, with reference to the following drawings in which:

Figure 1 shows schematically a computer system according to the invention;
Figure 2 shows a more detailed view of a computer device of Figure 1;
Figure 3 is a flow chart showing steps carried out at an initiating device;
Figure 4 is a flow chart showing steps carried out at a recipient device; and,
Figure 5 shows schematically three stages involved in the selection of devices for initialisation; and,
Figure 6 is a UML sequence diagram illustrating the communication and relationship between objects in the computer system.

In Figure 1, there is shown a computer system 10 which includes a plurality of computer devices 22, each with a respective software component 18 running thereon, the computer devices being interconnected by a plurality of data links 14. Tasks or jobs are dynamically allocated to the respective computer devices 22 by a server 16, such that the tasks or jobs allocated to one computer device depend on the tasks or jobs being performed by at least some of the other computer devices. In this way, the computer system 10 as a whole can

operate more efficiently than it would do in the absence of such coordination between the computer devices 22.

Each computer device is able to generate an initialisation parameter that is indicative of the need (or equivalently the urgency or desirability) for the computer device 22 to be initialised. When the initialisation parameter of a computer device 22 reaches a threshold value, that computer device acts as an initiating device 22a, and transmits a request message 20 over the data links 14 to the other computer devices 22, which devices then act as recipient devices 22b. The request message requests from the recipient devices 22b the current values of their respective initialisation parameters.

The initiating device 22a is configured to compare the received initialisation parameters from the recipient devices 22b together with its own initialisation parameter and to issue initialisation instructions to itself and/or one or more recipient device(s) 22b in dependence on the relative values of the initialisation parameters. Because the initialisation parameter of a computer device 22 is a measure of the need for that device to be initialised, the computer devices 22 most needing initialisation can be selected in favour of computer devices least requiring initialisation. Thus, those computer devices 22 which suffer disruption brought about by their initialisation can be chosen so as to reduce the overall disruption to operation of the system 10.

In effect, the recipient devices 22b participate in an "auction", the initialisation parameters representing "bids" for one or more available initialisation instructions. The initiating device 22a can be viewed as an "auctioneer" device, allocating initialisation instructions to the devices which place the highest bids: that is, those devices (including the initiating device) which are most in need of initialisation. Although in this example the bids are "sealed", in that the bid placed by one device is independent of bids placed by other devices (and different devices can "bid" at the same time), alternative embodiments are possible where a feedback routine is executed by the devices participating in the auction, such that the bid placed by one device is dependent on bids placed by another device.

The "auction" procedure or protocol can be summarised with reference to Figure 5 in terms of three stages labelled I to III. In stage I the initiating (auctioneer) device 22a announces its intention to hold an auction, by broadcasting a request message 11 to other recipient devices 22b. In stage II, devices not participating in another auction return bids

13 to the auctioneer device, whereas devices (indicated by shading) already participating in another auction with another auctioneer device (not shown) ignore the request message 11. The self-elected auctioneer device 22a also places a bid in its own auction. In stage III, after the auctioneer 22a has received and processed the bids, it informs the participants of the auction (non-shaded recipient devices 22b) of the "winning" devices which may be initialised (or equivalently another maintenance task can be initiated). To inform the participants, the auctioneer sends a message 15 to the participants, which message includes: the number of devices participating in the auction; the identity of the winning devices; and, the total number of devices participating in the auction.

Each computer device 22 may be housed in a respective housing or casing, the data links 14 being formed by cables which extend between the housings of the respective computer devices. In such a situation, the computer devices 22 may be situated in different geographical locations. Alternatively, the computer devices 22 may be located in a common housing, and even for example on the same circuit board.

A more detailed view of one of the computer devices 22 of Figure 1 is shown in Figure 2. The computer device 22 has a processor circuit 26 which includes a memory 31 and a processor 27 for respectively storing and executing the software component 28, as well as other software which may be running on the computer 22.

The software component 28 includes an operating system module 40 for controlling the operation of the computer device 22, and an application programme 42 for performing tasks. A monitoring stage 34 is provided for monitoring one or more status indicators indicative of the likelihood of a fault or potential fault with the computer device 22, and for producing an initialisation parameter in dependence on the monitored status data. The monitoring stage 34 may monitor the amount of memory space in the memory 31 that is deemed available by the operating system module 40. Thus, if the application programme 42 fails to release all the memory it has finished using, this "leaky programme" malfunction (or "bug") which could possibly lead to a fault can be monitored. When the amount of available memory drops below a threshold value, this can be classified by the monitoring stage as a fault or potential fault, and the monitoring stage can, in response to this fault or potential fault, issue an initialisation parameter indicating a high need for the operating system module 40 and/or the application 42 to be initialised.

However, in a preferred embodiment, the monitoring stage includes a timer 35 for measuring the elapsed time since the last initialisation (preferably the time since the last initialisation of the operating system module 40), the initialisation parameter being representative of this elapsed time. It is appreciated that because computer devices are likely to behave in a non-ideal way due to malfunctions such as the "leaky programme" malfunction, and because such malfunctions are likely to cause a deterioration in the operational capacity of a computer device over time, the elapsed time since the last initialisation can provide a useful measure of the need for a computer device to be initialised.

The monitoring stage may take into account the tasks being performed by the device on which it is running, more important tasks having a greater weighting on the initialisation parameter than tasks of a lesser importance. In one embodiment, the monitoring stage monitors whether the device on which it runs is performing one or more of a predetermined number of task, and will only issue a non-zero initialisation parameter to the if such a task is not being performed.

The initialisation may involve a re-boot (or equivalently a re-start) of the computer device 22, or alternatively the initialisation may simply require the application 42 to be killed and re-started, without re-booting the entire computer device 22. As a further alternative, the initialisation may require the amount of memory deemed in use by the application 42 to be re-set, for example to zero.

The initialisation parameter generated by the monitoring stage 34 is passed to an initialisation controller 32. The initialisation controller 32 is configured to determine whether the initialisation parameter (in this example, the elapsed time since the last initialisation) from the monitoring stage 34 is above a threshold value. If the initialisation parameter is above a threshold value, the computer device 22 then acts as an initiating device 22a as indicated in Figure 3, and executes an initialisation routine. In the initialisation routine, the initialisation controller 32 of the initiating device 22a transmits at an output stage 36 one or more requests for an initialisation parameter to the other computer devices 22 (acting as recipient devices 22b).

The initiating device 22a need not know the identity of the recipient devices, nor need the initiating device know the number of existing recipient devices. Therefore, the request for

initialisation parameters from the initiating device 22a will preferably be sent in the form of a broadcast message, normally on a well-known port, and normally without being addressed to any specified recipient devices 22b. The request message will preferably contain: a time stamp indicating the time at which the message was sent; the name, address, and/or other label which identifies the initiating device 22a from which the request message originates; information indicative of a port number on which a reply with an initialisation parameter should be sent; and a unique identifier for the request message.

For each computer device 22, there is provided a listening stage 36 configured to monitor data arriving at that device 22, so as to detect if a request for an initialisation parameter has been transmitted by another device 22 (the initiating device 22b), preferably on the chosen well-known port. When a request from an initiating device 22a is detected by the listening stage 36 of a recipient device 22b, the request is passed to the initialisation controller 32 of that device, in response to which the initialisation controller 32 executes a response routine (steps carried out by the recipient device 22b are shown in Figure 4).

As part of the response routine, the initialisation controller 32 of the recipient device 22b determines whether the recipient device 22b is involved in an initialisation procedure with another initiating device. If so, the initialisation controller 32 does not respond to the most recent request for an initialisation parameter. Otherwise, the initialisation controller 32 of the recipient device 22b will access status data or an initialisation parameter from the monitoring stage 34 of that device. The monitoring stage 34 determines if the recipient device 22b is busy performing one of a number of tasks, and if so, passes a busy signal to the initialisation controller 32. If the recipient device 22b is not busy, the monitoring stage 34 passes to the initialisation controller 32 an initialisation parameter indicative of the elapsed time since the last initialisation at the recipient device 22b. The initialisation controller 32 of the recipient device 22b then transmits the initialisation parameter to the initiating device 22a, unless a busy signal is received from the monitoring stage 34, in which case the initialisation controller 32 transmits the busy signal to the initiating device 22a in the form of an initialisation parameter which indicates the elapsed time since the last initialisation of the recipient device 22b is zero seconds. This busy signal will allow the initiating device 22a to treat the recipient device 22b as having the lowest possible priority for initialisation.

The listening stage 36 of the initiating device 22a captures any initialisation parameters received from recipient devices 22b in response to the request(s), and forwards these received initialisation parameters to the initialisation controller 32 of the initiating device 22a. The initialisation controller 32 then performs a comparison step using the initialisation parameters received from recipient devices 22b and the self-generated initialisation parameter from the monitoring stage 34 of the initiating device 22a. The comparison step involves ranking at least some of the initialisation parameters from different devices 22 in dependence on their respective magnitudes, and selecting a proportion (such as 10%) of the initialisation parameters having the highest values. An initialisation instruction is then sent to the respective computer devices 22 whose initialisation parameter falls within that proportion. (If one of those initialisation parameters used in the comparison steps originates from the initiating device 22b, the initialisation controller 32 of the initiating device 22b will issue a self-initialisation instruction to the operating system module 40 and/or the application 42 running on the initiating device). In this way, the devices 22 of the computer system 10 which most urgently require initialisation can be selected.

The comparison step may also involve taking into account the total number of responses to the request message, as well as the number of idle devices (i.e. devices which return a non-zero value initialisation parameter). Thus, 10% of idle devices may be selected for initialisation.

The number or percentage of devices chosen for initialisation may be set by an administrator or the server 16, and may depend on the predicted workload of the system (the workload may be predicted by analysing historical patterns in the way in which the server 16 allocates tasks to the respective device, or by inspecting a log of tasks awaiting allocation by the server 16). By predicting the likely workload on the system, it is possible to estimate how many devices or what percentage of devices can be re-booted without endangering the operation of the system as a whole. In this way, the re-boot times of the devices 22 are co-ordingated so as to make it likely that at any one time, there will be enough working devices to handle job requests arriving at the system 10.

The initialisation instructions from the initiating device 22a will preferably be sent as a common initialisation message to the recipients devices 22b, such that the recipient devices 22b receive the same message. The initialisation message will then contain information identifying which recipient devices 22b should be initialised, so that the

initialisation controller of an identified recipient devices 22b can act on the initialisation message. The initialisation message will preferably also contain information indicating the number of recipient devices 22b from which an initialisation parameter was received and the number of recipient devices 22b identified for initialisation.

The initialisation instruction for a recipient device 22b will preferably instruct the recipient device 22b to initialise within a time period. However, before instigating an initialisation, the initialisation controller 32 of the recipient device 22b will interrogate the monitoring stage 33 to determine if the recipient device 22b is performing any of a number of tasks (and optionally whether the tasks(s) will be continued beyond the time period allowed for in the initialisation request). If so, the recipient device 22b is deemed to be busy, and the initialisation controller of the recipient device 22b ignores the initialisation instruction. Optionally, the initialisation controller 32 of the recipient device 22b can be configured to carry out an initialisation in response to the initialisation instruction if the elapsed time since the last initialisation exceeds a threshold value, regardless of whether the recipient device 22b is busy or idle.

In Figure 6 there is shown a UML sequence diagram illustrating the communication and relationship between software objects in the computer system (each object or process is represented as a box in Figure 6). Each device 22 that can potentially take part in the auction runs a listening process 61 indicated by the RunListener time line. This process waits for communication from other devices. When a device has reached the threshold necessary to hold an auction, a process 63 indicated by the RunAuction time line starts. This RunAuction process 63 creates a BidManager process 65 and informs it which communication port to use. The BidManager process 65 creates a BidCollector process 67 and passes it details of the communications port on which to receive bids. The BidManager process 65 contacts the RunListener running in each devices by sending a message 69. When the RunListener process 61 of a device 22 receives such a message 69 it creates a BidMaker process 71 for that device supplying device name and communications port number. The BidMaker process 71 sends a placeBid message 73 to the BidCollector 67. The BidManager also creates a BidEvaluator process 75 which calculates the winning bids i.e., which selects devices for initialisation. Winners (i.e., devices selected for initialisation) are informed via the BidCollector process and the BidMaker process. The sequence shown in Figure 6 is preferably implemented using (or written in) the "Java" language.

The embodiment described above is suitable for a collection of computers. Because the protocol in this embodiment does not need to know the membership details of the collection it is able to continue to work in the face of partition of the collection. As an example of this, if we consider 2 islands of computers, for example 50 in each, linked by a network connection into a single pool where the "auction protocol" of the invention is active, scheduling of maintenance or initialisation will take place as and when required. Should the network connection fail, each island can continue scheduling maintenance within itself, until the connection is restored when the whole pool can participate again.

In another embodiment (not shown), the software components 28 run on a common computer device under the control of a common operating system. An example of this is a server running different programs for different users. In such a situation, the software components 28 will not normally include the operating system, such that the operating system can continue to run even when one or more of the components 28 are initialised.

As can be seen from the above description, the present invention will be particularly useful in a system having components which work together as a group and which each require initialisation at time intervals, allowing components to initialise without excessively disturbing the group dynamics of the system.